Reconstruction of the Z boson signal in $Z \to \mu^-\mu^+$ and $Z \to b\bar{b}$ decays

Marino Missiroli
CDF - Padova group

September 23, 2011

Abstract

A preliminary study for the signal extraction of a Z boson decaying either in leptons and hadronic jets is presented. For the leptonic decay we consider muon-antimuon production in a CDF data sample (period 18) acquired with the MUON_CMUP8_DPS trigger. For the hadronic case, the analysis focuses on the $Z \to b\bar{b}$ decay mode in order to verify the possibility to extract this signal from a CDF dataset (periods 20,21 and 22) selected by the DIJET_BTAG trigger system.

Contents

1 Leptonic decay $Z \to \mu^- \mu^+$

As a introductory exercise, we consider the leptonic decay of the Z boson in a muon-antimuon pair.

The upside in studying this decay mode is the limited amount of background, due to the features of muons, which are easy to identify: as a matter of fact, in particle physics experiments, processes with leptonic or semileptonic signatures are often the golden channels for the analysis of a certain signal.

The goal is to reconstruct the invariant mass of the this decay selecting high-P_t muons: the following analysis shows that standard conditions for muon identification and minimal constraints on the kinematic proprieties of $\mu^- \mu^+$ pairs are sufficient to extract the Z signal.

We make use of the data sample collected by the CDF detector during the period 18 data acquisition with the MUON_CMUP8_DPS trigger.

The Pythia Monte Carlo (MC) program is used to generate the inclusive $Z \to \mu^- \mu^+$ process, with $\sigma = 355 \pm 3$ pb and $M_{inv} > 20$ GeV.

For the analysis of our data samples, we basically use the main cuts selected in CDF note 8262 [1], which is our primary reference for this matter; in the following we outline a brief description of these constraints, comparing MC and data results.

1.1 Muon selection conditions

The cuts used for muon selection are summarized below.

I1	CMU stub = 1	E1	$P_t > 20$ GeV	G1	$\lvert\eta_\mu\rvert < 1$
I2	CMP stub = 1	E2	$E_{em}^{cal} < 2$ GeV	G2	$\lvert z_{0\mu}\rvert < 60$ cm
I3	$\lvert\Delta x_{CMU}\rvert < 3$ cm	E3	$E_{had}^{cal} < 6$ GeV	G3	$\lvert z_{vertex}\rvert < 60$ cm
I4	$\lvert\Delta x_{CMP}\rvert < 5$ cm	E4	$E_t^{iso04}/\ P_t < 0.1$	G4	$\lvert z_\mu - z_{vertex}\rvert < 5$ cm
				G5	$\lvert d_0\rvert < 100\ \mu$m

Table 1 Identification (I), energetic (E) and geometric (G) cuts for muon selection

The CMUP stubs (I1-2) are requested to confirm track detection in both the central muon chambers; (I3-4) are standard geometric cuts related to these detectors. These four cuts are the ones that should provide the best discrimination between muon tracks and other different particles.

(E1) is a standard cut for the minimum transverse energy of a high-P_t muon; (E2) and (E3) are upper limits for the energy release in the electromagnetic and hadronic calorimeters and are necessary to distinguish muon,

from other particles, like pions, originating from underlying events or other particles decays; (E4) is a condition to characterize muons from particles with greater energy loss.

The (G2-G3) constraints are the standard ones for track and vertex positions, and their difference; (G5) is an upper limit on the track impact parameter relative to the primary vertex of the event. Unlike [1], we add the (G1) condition (central muons) to be consistent with the choice of selecting only the tracks which are detected by the CMU and the CMP; these detectors, in fact, cover only the $|\eta| < 1$ portion of geometric space in the detector, so selecting tracks with $|\eta| > 1$ for muon ID is basically pointless in this framework. As a matter of fact, once all the other restrictions listed in Table 1 are satisfied, the (G1) condition is automatically satisfied as well.

Since our goal is reconstruct the $Z \to \mu^-\mu^+$ signal, we are interested in events with at least *two* tracks satistying the above selection.

Table 2 provides MC efficiencies ϵ calculated as the ratio between the number of events with at least two tracks satisfying a certain set of conditions (identification, energetic, geometric) and the total number of events generated in the MC simulation.

	ϵ
I / total	21.8 ± 0.2 %
I+E / total	13.7 ± 0.2 %
I+E+G / total	11.6 ± 0.2 %

Table 2 Cut efficiencies, with respect to the total number of generated events in $Z \to \mu^-\mu^+$ MC, for events with at least two tracks satisfying identification conditions (I), identification and energetic conditions (I+E), identification, energetic and geometric conditions (I+E+G)

These values confirm that the cuts involving the CMU and CMP detectors are clearly the most discrimant in muon identification. After the application of all identification, energetic and geometric constraints we find that every MC generated event has *at most* two tracks passing all the cuts, so with the conditions of table 1 we're able to identify at most *one* pair of muons for every MC event; in the following we refer to the events in which this pair is present as "two muons events" (2μEv).

1.2 $\mu^-\mu^+$-pair selection for the $Z \to \mu^-\mu^+$ signal

We apply on "two muons events" further requirements to identify muon-antimuon pairs associated to a Z decay; these cuts are listed in table 3.

P1	$Q_{\mu 1} \cdot Q_{\mu 2} < 0$
P2	$\lvert z_{\mu 1} - z_{\mu 2} \rvert < 4$ cm
P3	$71 < M_{inv}(\mu 1, \mu 2)$ (GeV) < 111

Table 3 Muon-antimuon pair (P) cuts for $Z \to \mu^- \mu^+$ events selection

(P1) is the opposite-charges condition. (P2) is the geometric restriction applied in [1] to reject cosmic ray tracks and ensure that both muons are coming from the same vertex. (P3) is actually different from the corresponding cut in [1], where a [81,101] GeV interval is selected. Data analysis shows that we can use a wider range and still get good results.

We report the efficiencies obtained from the $Z \to \mu^- \mu^+$ MC simulation for the three cuts described above; here, by efficiency we mean the ratio between the number of "two muons events" passing a certain condition and the total number of "two muons events".

	ϵ
P1 / 2μEv	$100 - 0.2$ %
P1+P2 / 2μEv	$100 - 0.2$ %
P1+P2+P3 / 2μEv	91.1 ± 2.0 %

Table 4 Cuts efficiencies for $\mu^- \mu^+$ pair conditions

It's worth mentioning the fact that in every event of the MC sample there are at most two tracks passing the cuts of table 1 and, in these cases, the two muons always have opposite charges and small relative distance on z-axis, i.e. they are produced in the same vertex.

Figure 1 shows the distributions of some of the most important variables, considering only the tracks which satisfy all the conditions described above for muon and pair selection. MC and DATA histograms are plotted after being normalized to enable a qualitative comparison.

These plots confirm that our MC simulation for the $Z \to \mu^- \mu^+$ process correctly reproduces the signal we're extracting from the experimental dataset. This is what we expected, since we already know that the cut selection we're using is the optimized one for this kind of signal.

Given the number of events with a $\mu^- \mu^+$ pair whose invariant mass is in the 71-111 GeV range in the MC sample, the branching ratio of the $p\bar{p} \to Z \to \mu^- \mu^+$ process and the integrated luminosity of the analyzed dataset (CDF period 18) $\mathcal{L}_{18}$, it's possible to estimate the number of Z-events we expect to find in the data sample. As cut efficiency ϵ_{cut} we take the ratio

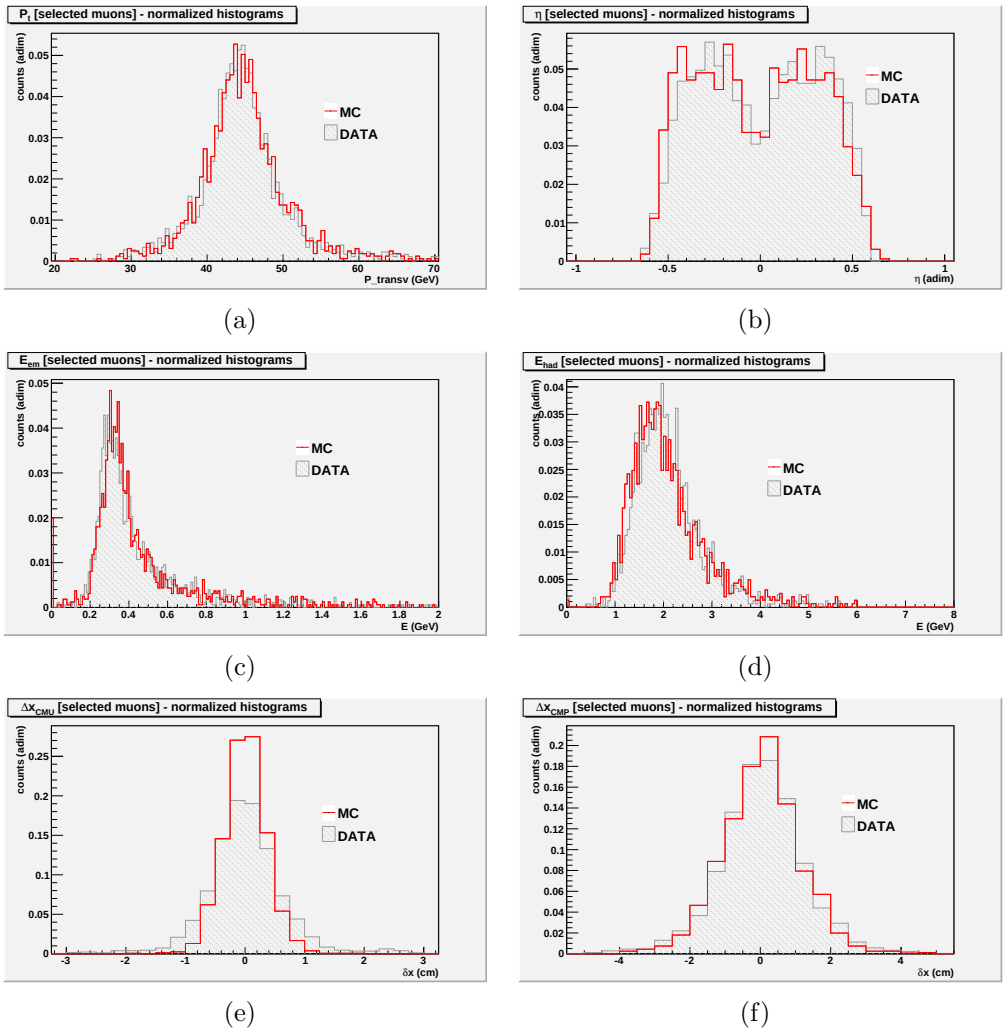


Figure 1 P_t (a), η (b), E_{em} (c), E_{had} (d), Δx_{CMU} (e) and Δx_{CMP} (f) normalized distributions for muons satisfying all the conditions of table 1 and 3. MC simulation and experimental DATA outcomes are compared

between the number of events satisfying all the above cuts and the total number of events in the MC simulation. For the MUON_CMUP8_DPS trigger we have an estimated efficiency of 95 % and we use this value as trigger efficiency $\epsilon_{trigger}$ in our computation.

$$\sigma = \sigma(p\bar{p} \to Z) \cdot BR(Z \to \mu^-\mu^+) = 355 \pm 3\,\text{pb} \qquad \mathcal{L}_{18} \simeq 264\,\text{pb}^{-1}$$

$$\epsilon_{trigger} = 0.95 \qquad \epsilon_{cut} = 0.011 \pm 0.001$$

So, the estimated number of Z-events in the analyzed dataset is

$$N_{data} = \sigma \cdot \mathcal{L}_{18} \cdot \epsilon_{trigger} \cdot \epsilon_{cut}^{MC} = 979 \pm 89$$

where the error is calculated just propagating the σ and ϵ_{cut} errors.

We apply the same muons and muon-antimuon pairs selection constraints to the experimental data at our disposal and find that the number of events with at least a $\mu^-\mu^+$ pair, whose invariant mass is in the 71-111 GeV range, is $N_Z = 886$ on 15639717 triggered events; this is compatible with the estimate made using the MC cut effinciency.

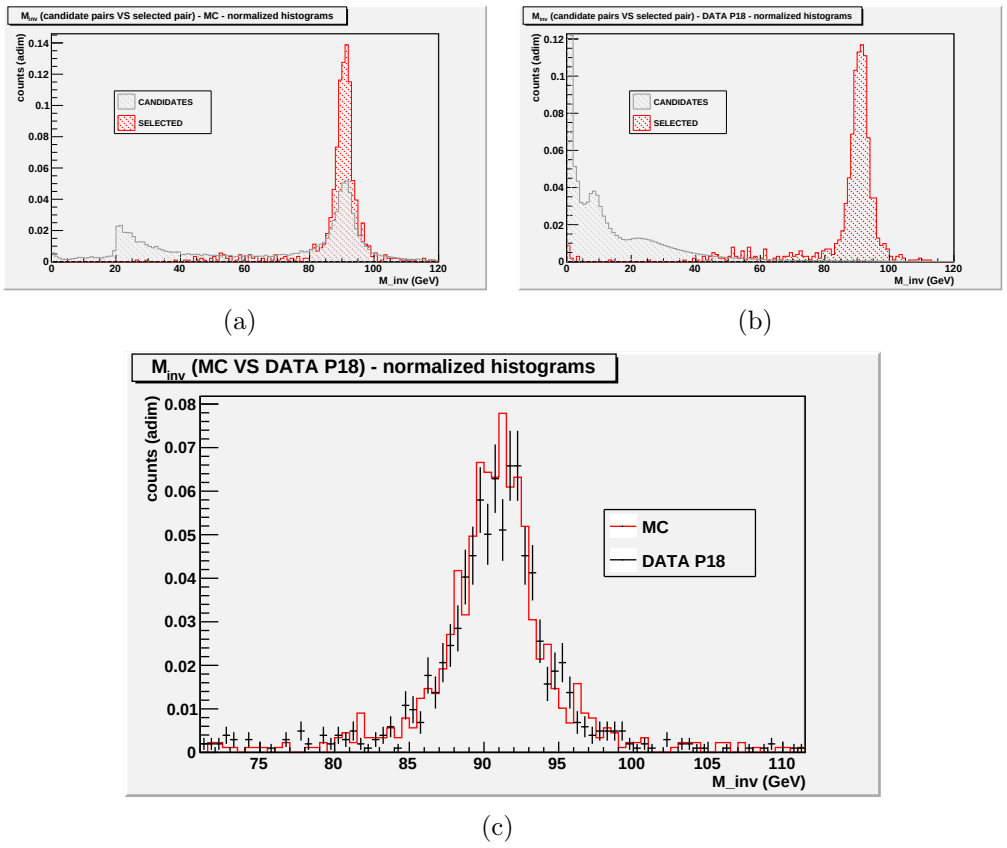


Figure 2 Invariant mass distribution for all candidate pairs and for the selected one in (a) MC simulation and (b) PERIOD 18-DATA. (c) M_{inv} distribution post-selection in MC and DATA.

Moreover, with all cuts applied, we are able to select (at most) one $\mu^-\mu^+$ pair [1] for every event in the dataset as well as in the MC sample.

This allows us to reconstruct the invariant mass of the $Z \to \mu^-\mu^+$ decay mode ($71 < M_{inv}(\mu^-, \mu^+)$ (GeV) < 111). The invariant mass distributions of all candidate pairs (pre-selection)[2] and of the single pair obtained post-

[1]Due to the Z production cross-section in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV, is nearly impossible to have more than one Z produced per event.

[2]The only condition applied to the so called "candidate" pairs is (P1) (charge).

selection are compared in Figure 2(a) for the MC simulation and in Figure 2(b) for the PERIOD 18 data sample. Figure 2(c) shows the invariant mass distribution of the $\mu^-\mu^+$ (post-selection) for MC and DATA. All histograms are normalized to unity.

These plots show that the present analysis allows for a clear reconstruction of the $Z \to \mu^-\mu^+$ signal.

1.3 Final results

The last step to complete this introductory study is to determine the peak of the invariant mass distribution obtained from the our data sample applying all the conditions described above: within the precision bounds of this analysis, this value should correspond to the Z boson mass, i.e. $M_Z = 91.19$ GeV.

Figure 2(c) shows that in the 71-111 GeV interval there is a limited and almost homogeneous amount of background, so trying to fit the background signal with a first-order polynomial is a reasonable procedure. Supposing the Z signal is compatible with a gaussian distribution, we use the following function to fit the M_{inv} distribution in the 71-111 GeV interval

$$f(x) = C \exp\left(\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2\right) + A + Bx \tag{1}$$

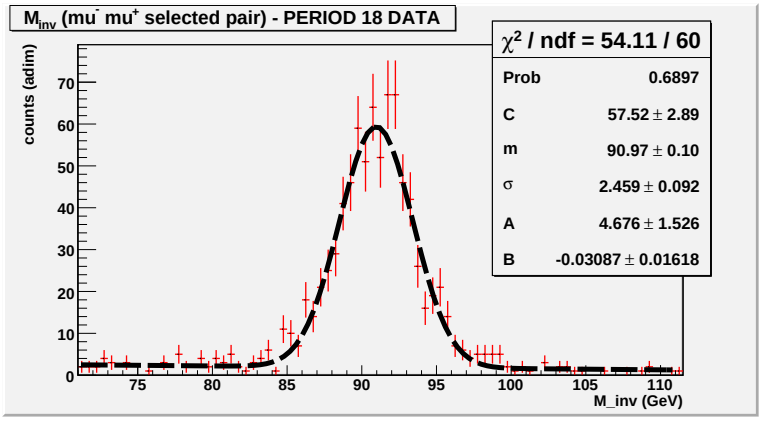


Figure 3 M_{inv} distribution (1 GeV per bin) for the selected $\mu^-\mu^+$ pair (points) and corresponding fitting function (line)

$$\left|\frac{m - M_Z}{\sigma_m}\right| = \left|\frac{90.97 - 91.19}{0.10}\right| = 2.2 \tag{2}$$

χ^2 / *ndf*	54.1 / 60
Prob	0.60
C (adim.)	57.5 ± 2.9
m (GeV)	91.0 ± 0.1
σ (GeV)	2.5 ± 0.1
A (adim.)	4.7 ± 1.5
B (GeV^{-1})	(-3.1 ± 1.6)· 10^{-2}

The goodness of the χ^2-test confirms our first assumptions on the fitting function. The mean value of the distribution is 91.0 ± 0.1 GeV; we accept this as a good estimate of the Z mass (91.19 GeV), considering that the fit is binned and doesn't take into account any kind of systematic error, but only statistical ones.

In conclusion, throughout this "pilot" analysis of the leptonic decay channel $Z \to \mu^-\mu^+$, all the main characteristics of the process, from basic kinematic distributions to the decay invariant mass computation, have been verified and the corresponding results are pretty good.

2 Hadronic decay $Z \to b\bar{b}$

We proceed with the study of the Z boson hadronic decay mode into b-jets. The analysis of this process is far more complicated than the previous one, due to additional difficulties in dealing with hadronic jets and extracting the Z signal from the overall data sample. On the other hand, this channel is one of the best to extract information on b-jets features, which can then be applied to other searches involving this type of jets. These features are mainly

1. the b-jet energy scale factor, i.e. the necessary correction to account for the systematic error in the jet energy measurement, due to the imperfect knowledge of the fragmentation proprieties of b-quarks (hadronization)

2. the b-jet energy resolution, which is basically the experimental accuracy in b-jet energy measurements

Both these quantities can account for correcting the main source of uncertainty in precision measurements of the *top* quark mass and can be exploited as a tuning tool in searches for a low mass Higgs boson ($114 < M_H < 135$

GeV) since $H \to b\bar{b}$ decay mode is the dominant one in this mass range, which hasn't been excluded yet by LHC results. Knowing the Z mass, generally measured using its leptonic decays, the $Z \to b\bar{b}$ signal provides a way to evaluate these quantities.

The main problem in performing this analysis is the reduction of the number of background events due to *bottom*, *charm* and light quarks production in QCD processes, which constitutes the greater part of the overall signal. Without any restriction applied, the Z-events contribution is very small compared to the whole $b\bar{b}$ production since

$$\sigma(p\bar{p} \to Z \to b\bar{b}) \sim \text{nb} \qquad \sigma(p\bar{p} \to b\bar{b}) \sim \mu\text{b}$$

An off-line selection is then necessary to make the Z fraction emerge with respect to the background.

The Pythia program is used to generate MC simulations of the individual processes involved, in order to study what are the kinematical proprieties of each of them and how they contribute to the composition of the complete signal. Starting from these models we can study the best set of constraints to highlight the Z production process and then apply the same cuts selection to the data sample in order to actually extract the $Z \to b\bar{b}$ signal.

In this preliminary study, we use part of CDF experimental data in Run-II, produced by $p\bar{p}$ collisions with a 1.96 TeV center-of-mass energy. Data acquisition is performed using the DIJET_BTAG trigger system. For details on the CDF detector and the DIJET_BTAG trigger see [3] and [4].

The Pythia generator is used to reproduce the following processes:

- $Z \to b\bar{b}$ ($M_{inv} > 30$ GeV)
- QCD $b\bar{b}$ (OLD version, P_t >10 GeV)
- QCD $c\bar{c}$ (OLD version, P_t >12 GeV)

In the "OLD" versions of MC simulations quark production is achieved only through the flavour creation process.

After a brief discussion of the necessary requests for b-jets off-line reconstruction, the analysis framework is the following. First, we follow the event selection described in [5], in which the $Z \to b\bar{b}$ signal is successfully selected considering only events with two back-to-back jets in the central calorimeter region; we refer to this approch as the *2-jets analysis*. After that, we introduce and motivate a possible new path of analysis, in which the goal of reducing the background signal is pursued through the selection of events with a third well reconstructed jet with appropriate characteristics; we refer to this approch as the *3-jets analysis*.

2.1 Event reconstruction

In this subsection we describe the basic choices for the reconstruction of events with b-jets in the final state; these are the same in the *2-jets* and *3-jets* analyses.

ENERGY CORRECTION

The standard CDF jet energy correction package is used to associate to each jet the energy of the originating parton, applying several levels of correction; see [6] for details. We follow the standard choice made in CDF analyses, which is to use jet energies corrected up to level 5; higher levels of corrections are process-dependent and are excluded to maintain the outcomes of this study as general as possible.

CONE SIZE

Another important characteristic regarding jet reconstruction is the jet cone size $R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$. The two standard values for this parameter are $R = 0.4$ and $R = 0.7$. For a detailed study of jet cone size see [7]. In this analysis we consider only "cone 07" jets for two main reasons:

1. the DIJET_BTAG trigger implements the cone algorithm considering only jets with $R = 0.7$

2. a preliminary study of the invariant mass distribution, using the two E_t(L5) leading jets, with MC simulations shows that "cone 07" jets provide a better reconstruction of the Z signal, with a ~ 90 GeV peak. On the other hand, the same distribution with "cone 04" jets peaks at a lower energy value: this is not only different from what we expect for this kind of simulation, but also disadvantageous in view of the attempt to extract the Z signal from the dominant QCD background, since we already know that the dijet invariant mass distribution for the QCD $b\bar{b}$ template peaks below 90 GeV; so, using "cone 04" jets, the discrimination between signal and background would also be harder.

b-TAGGING ALGORITHM

One important feature of the bottom quark is its lifetime, which is longer than that of the other quarks; a b-quark travels, indeed, a few millimeters before hadronising, so selecting jets whose cone contains a secondary decay vertex is the most efficient method to identify b-jets. This is done using the SecVtx b-tagging algorithm, which basically tries to reconstruct a common decay vertex for multiple charged tracks associated to the same jet; see [8]

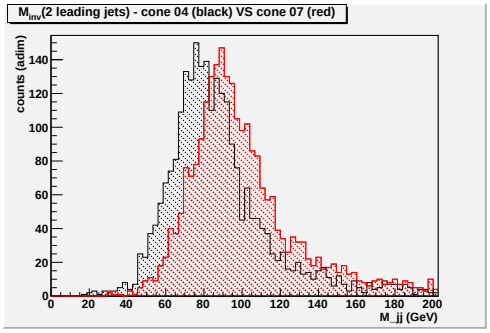


Figure 4 M_{inv} distribution for the two E_t(L5)-leading jets (both with $E_t(L5) > 15$ GeV), reconstructed with cone size $R = 0.4$ (black) and $R = 0.7$ (red)

for details. A jet is *b-tagged* if the algorithm reconstructs a secondary vertex inside the jet cone size. If the secondary vertex is found to be outside the cone, the jet is said to be *negative b-tagged*; the latter subset contains light quarks and gluon jets and provides a sample to be used to estimate the algorithm mis-tagging rate.

2.2 The *2-jets* analysis

2.2 Introduction

As anticipated, the first part of this section is a reproduction of the analysis conducted in [5], which is briefly described as follows.

In the paper cited above, the "signal over background" ratio (S / B) is optimized considering only b-tagged jets emitted in opposite directions and in the central region, exploiting the lower probability of QCD initial and final state radiation in Z-production events. With these selection, the background template is composed mainly by $b\bar{b}$ quarks produced via flavour creation and only low-momentum Z bosons are selected as signal. This analysis was performed on a CDF dataset acquired with the Z_BB trigger system. In this paper, as already mentioned, we always analyze a CDF data sample collected with the new DIJET_BTAG trigger system, even when we reproduce the event selection described in [5].

2.2 Sample composition

A preliminary step towards the extraction of Z signal is verifying the flavour content of the dataset in our hands, i.e. determining the percentages of *bottom*, *charm* and *light quarks* in the DIJET_BTAG data sample. This is done through the reconstruction of the invariant mass of the charged tracks inside the jet, the so called vertex mass, which is directly linked to the rest mass of the quark originating the jet; the reconstruction of the vertex mass is performed on tagged jets in the dataset and in MC simulations modeling the single QCD $b\bar{b}$ and $c\bar{c}$ contributions; as for the light quark template, our MC simulations don't provide enough statistics to create a decent template, so we choose a different, often used, approch, consisting in the reconstruction of the light quark contribution directly from the dataset, selecting anti-tagged jets instead of tagged ones to determine the mis-tagging contribution of the SecVtx algorithm [3]. The class `TFranctionFitter` of ROOT is used to perform the calculation of the percentages for the different templates.

In this subsection we accomplish this using the event selection described in [5]; so, since we deal with the 2-jets analysis, for the b and c quark template we can use the OLD MC simulations, which reproduce only the flavour creation process.

The event selection is determined by the following conditions

1. NO jets in the frontal region,
   i.e. NO jets with $E_t(L5) > 10$ GeV && $|\eta| > 1$

2. AT LEAST two jets in the central calorimeter,
   i.e. with $E_t(L5) > 20$ GeV and $|\eta| < 1$

In these selected events we apply the SecVtx tagging algorithm only on the two E_t(L5)-leading jets; from here to the end of this subsection only these two jets are considered.

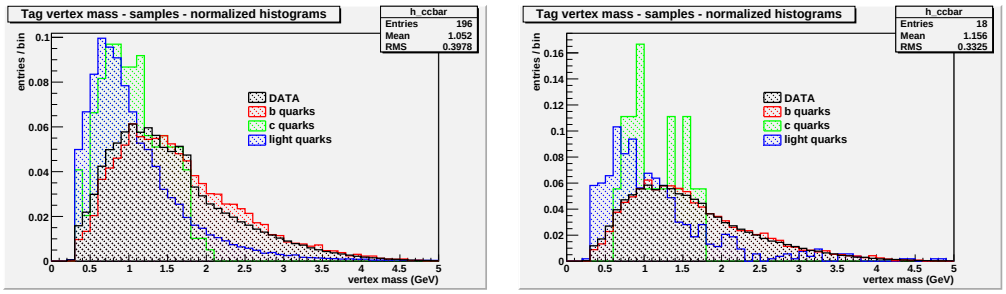
We perform the sample composition on two subsets of events

- *single tag events* : at least one of the two leading jets is tagged

- *double tag events* : both leading jets are tagged

In both cases, the vertex mass template is constructed with every tagged E_t(L5)-leading or subleading jet in the event. Obviously, double tag events are a subset of single tag events.

[3] Since light quark/gluon jets are not characterized by the presence of a secondary vertex, their contribution is accounted for in the mis-tagging template



(a) (b)

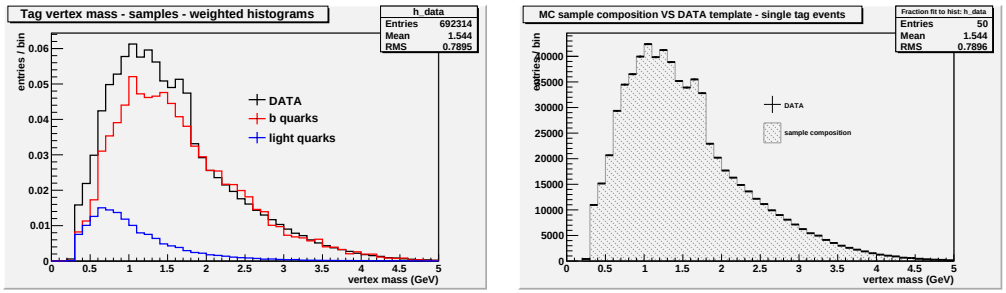
Figure 5 Data (black), b (red), c (green) and light quark (blue) templates, normalized to unity, in (a) single tag events and (b) double tag events

Figure 5 shows the superposition of the four templates, all normalized to unity, both in single tag events and in double tag events; just by looking at the figure, it's clear that the c template is not well reconstructed even in the single tag events case, due to insufficient statistics. The fit procedure doesn't converge if this template is included. Anyway, this problem can be easily overcome, observing that the c contribution is very similar to the light quark one, so descarding the former has the only effect of increasing the percentage of the latter, but, since what we really want to estimate is the b quark contribution, this sacrifice doesn't compromise the goodness of our analysis in a first approximation. Moreover, in this case a precise calculation is not necessary: we just need to verify that our data sample is mainly formed by b quarks, especially in double tag events.

We proceed without the c contribution and compute the b and light quark fractions in the data template for single and double tag events.

The results of the fit procedure totally respect our expectations: in both cases the b fraction is by the dominant one and in double tag events it accounts for over 90 % of the data template. This confirms that selecting events in which both leading jets are b-tagged provides us with a sample with the highest $b\bar{b}$ purity; in this case, we can then consider the QCD $b\bar{b}$ production process as the only component of the background signal. Finally, we note that the χ^2 value is too high in single tag events and too low in double tag events; in the first case, every template is characterized by good statistics and the result is not good enough, probably because of the c template absence; in the second case, at least the light quark template has few entries, so the errors are overestimated and the χ^2 value is too low. Anyway, as already mentioned, the sample composition is a preliminary check to the actual analysis and it's not intended to be a precision measurement, but rather a qualitative one. According to the expectations and to similar

studies performed in other analyses, we can definitely say that our results are good.

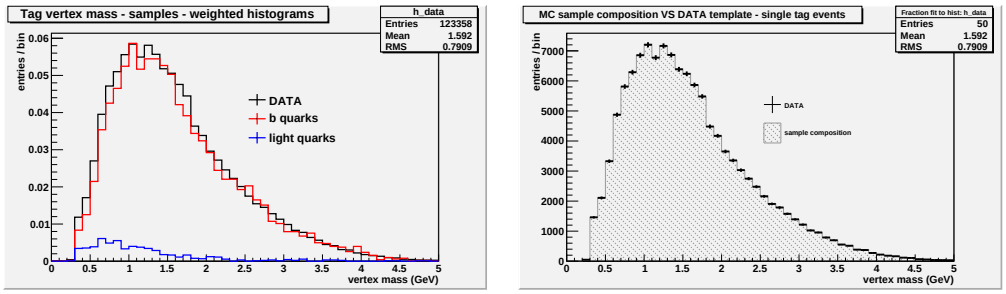


(a) (b)

Figure 6 Single tag events: (a) Data (black), b (red) and light quark (blue) templates, weighted according to the fit results. (b) Data template (bars) and sample composition outcome (filled histogram)

$b\bar{b}$	84.9 ± 1.2 %	χ^2 / ndf	86.3 / 48
$q\bar{q}$	15.1 ± 0.9 %	Prob	0.0006

Table 5 Templates percentages and χ^2 test for the sample composition fit in single tag events



(a) (b)

Figure 7 Double tag events: (a) Data (black), b (red) and light quark (blue) templates, weighted according to the fit results. (b) Data template (bars) and sample composition outcome (filled histogram)

$b\bar{b}$	94.1 ± 2.0 %	χ^2 / *ndf*	36.8 / 48
$q\bar{q}$	5.9 ± 1.4 %	Prob	0.88

Table 6 Templates percentages and χ^2 test for the sample composition fit in double tag events

2.2 Optimized cuts selection

We proceed applying the events selection described in [5]; first, we consider MC simulations reproducing the $Z \to b\bar{b}$ signal and the QCD $b\bar{b}$ background separately, comparing the effect of this selection on the two different processes, and then we apply the same cuts to the data sample.

In [5] the optimization of the cuts selection brings to the following set of conditions.

1. $E_t(L5)$(leading jet)[4] > 22 GeV
2. $E_t(L5)$(subleading jet) > 22 GeV
3. $|\eta|$(leading jet) < 1
4. $|\eta|$(subleading jet) < 1
5. b-tag(leading jet) $= 1$
6. b-tag(subleading jet) $= 1$
7. $3 < \Delta\phi_{12}\,(\mathrm{rad}) < 3.3$ (back-to-back leading jets)
8. $E_t(L5)$(3rd jet) < 15 GeV

In this subsection we apply this constraints to our dataset and study the corresponding cut efficiencies in MC simulations. We use a 3-level off-line simulation of the DIJET_BTAG trigger which also accounts for the scale factor related to the calorimeter efficiency (`calorimeter_turnon` function); the events that pass the trigger selection are labeled as "triggered" in the following.

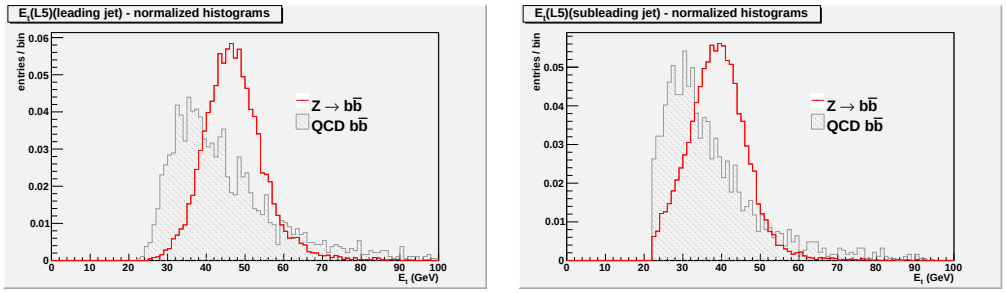
In figure 8 the transverse energy of the two leading jets are pictured for both MC simulations, Z signal and $b\bar{b}$ background, after the whole cut selection described earlier in this subsection.

[4]The jets list is ordered by decreasing values of E_t(L5)

ϵ	$Z \to b\bar{b}$	QCD $b\bar{b}$
triggered / total	4.46 ± 0.02 %	0.23 ± 0.08 %

ϵ	$Z \to b\bar{b}$	QCD $b\bar{b}$
1.+2. / triggered	98.6 ± 0.1 %	96.1 ± 0.3 %
1.+2.+3.+4. / triggered	76.7 ± 0.2 %	80.6 ± 0.6 %
1.+2.+3.+4.+5.+6. / triggered	19.6 ± 0.2 %	24.3 ± 0.6 %
1.+2.+3.+4.+5.+6.+7. / triggered	11.3 ± 0.2 %	12.1 ± 0.5 %
1.+2.+3.+4.+5.+6.+7.+8. / triggered	8.2 ± 0.1 %	10.0 ± 0.4 %

Table 7 Trigger simulation and cuts efficiencies ϵ in $Z \to b\bar{b}$ and QCD $b\bar{b}$ MC samples, corresponding S / B ratio and significance S / $\sqrt{S+B}$, relative to the events selection described in [5] and above



(a) (b)

Figure 8 E_t(L5) of (a) leading and (b) subleading jet in $Z \to b\bar{b}$ and QCD $b\bar{b}$ MC simulations after the events selection

Figure 9 provides a comparison between $Z \to b\bar{b}$ MC and QCD $b\bar{b}$ MC for the invariant mass reconstructed with the two E_t(L5)-leading jets. As expected, the QCD distribution peaks at an energy value lower than 90 GeV and its shape is mainly determined by the minimum energy cut implemented in the event selection, since no resonances are produced in this type of process.

Just like we did in the analysis of the Z leptonic decay, we can now get an estimate of the expected number of Z-events in our data sample using the cut efficiency calculated from the MC simulation, which is the ratio between the number of selected events and the number of generated events. For the

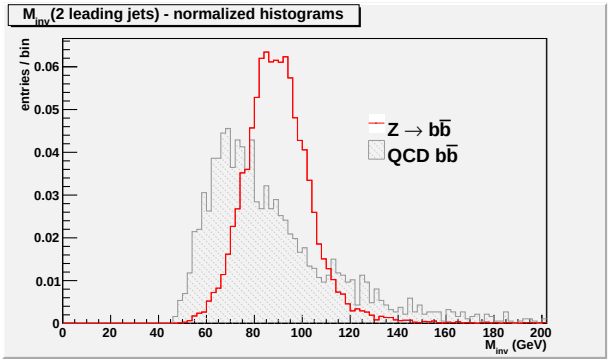


Figure 9 M_{inv} distribution for the two E_t(L5)-leading in $Z \to b\bar{b}$ and QCD $b\bar{b}$ MC simulations after the events selection

DIJET_BTAG trigger $\epsilon_{trigger}$ we consider an approximate efficiency equal to 95 %. In this case it's also necessary to consider the SecVtx tag efficiency ϵ_{tag} for every b-tag requested in the events selection, i.e. two in our case; a reasonable estimate for this value is 90 %. From the following values

$$\sigma = \sigma(p\bar{p} \to Z) \cdot BR(Z \to b\bar{b}) = 1129 \pm 22\,\text{pb (NLO)} \qquad \mathcal{L}_{20} + \mathcal{L}_{22} \simeq 440\,\text{pb}^{-1}$$

$$\epsilon_{trigger} = 0.95 \qquad \epsilon_{tag} = 0.90 \qquad \epsilon_{cut}^{MC} = 0.00365 \pm 0.00001$$

the estimated number of Z-events in CDF period 20 and 22 data sample is

$$N_{data} = \sigma \cdot (\mathcal{L}_{20} + \mathcal{L}_{22}) \cdot \epsilon_{trigger} \cdot (\epsilon_{tag})^2 \cdot \epsilon_{cut}^{MC} = 1395 \pm 27$$

where the error is computed just propagating the σ and ϵ_{cut}^{MC} errors.

2.2 Invariant mass distributions and preliminary signal extraction

The next step is the application of the events selection described above to the CDF dataset; this is the sample from which the Z signal can be extracted. In order to do this, the following strategy is used: the invariant mass distributions for the two E_t leading jets in DATA, MC of QCD $b\bar{b}$ and MC of $Z \to b\bar{b}$ are computed, using the MC simulations to model the different contribution to the overall signal; the same algorithm used in the sample composition fit is applied to these three templates in the attempt of estimating the percentage of events for the two processes in the analyzed dataset.

This can't be considered as a precise measurement, because a series of systematic errors and the corresponding corrections should be taken into account to make our analysis thorough. What we can obtain from a similar

procedure is to estimate the fraction of Z-events in the data sample considered and verify its compatibility with the same value calculated from the MC efficiency; in conclusion, we can't rigorously reconstruct the Z-signal this way, but we can at least figure out if the signal can be extracted or not perfoming an approximate study like the one described.

This signal composition procedure is performed with a 2 GeV binning, selecting an invariant mass interval between 50 and 200 GeV.

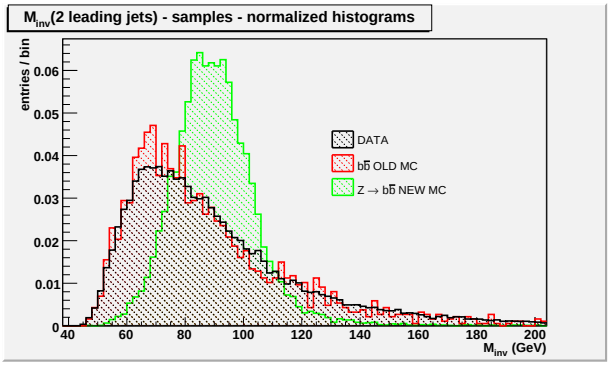
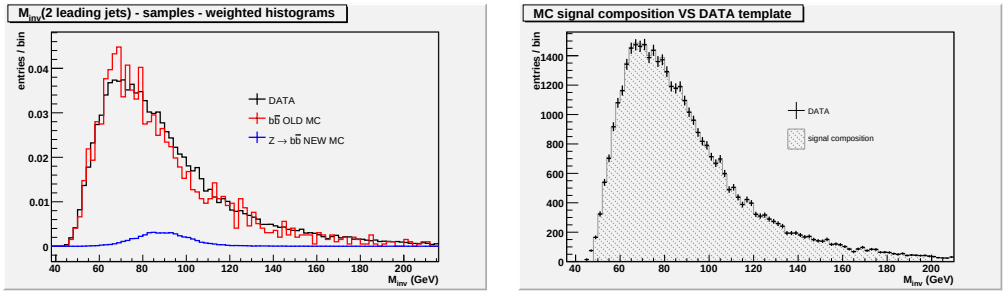


Figure 10 M_{inv} distributions for CDF period 20 and 22 data (black), QCD $b\bar{b}$ MC (red) and $Z \to b\bar{b}$ MC (blue) after the events selection



(a) (b)

Figure 11 (a) Data (black), QCD $b\bar{b}$ MC (red) and $Z \to b\bar{b}$ MC (blue) invariant mass distributions, weighted according to the fit results.
(b) Data distribution (bars) and signal composition outcome (filled histogram)

QCD $b\bar{b}$	95.1 ± 3.0 %	χ^2 / ndf	78.4 / 74
$Z \to b\bar{b}$	4.9 ± 2.4 %	Prob	0.34

Table 8 Templates percentages and χ^2 test for the signal composition fit

This procedure estimates a 4.9 ± 2.4 % contribution from $Z \to b\bar{b}$ events in the DIJET_BTAG data sample.

Given the number of entries in the M_{inv} data template, the number of Z-events predicted by the fit is

$$N_{data}^{fit} = 39625 \cdot (0.049 \pm 0.024) = 1931 \pm 948$$

This value is compatible with the prediction obtained using the MC efficiency; anyway, the greatness of the error forces us to consider this just as a qualitative check for our analysis.

Even though the result can't be said to be precise, this study highlights at least the possibility of extracting the Z signal from the DIJET_BTAG dataset, confirming our expectations and also providing a first estimate for the Z-events fraction using the *2-jets* analysis selection.

2.3 The *3-jets* analysis

2.3 Introduction and motivation

In this last subsection, we propose a new analysis framework, radically different from the one used in [5]. From preliminary studies of MC simulations we are able to verify that some of the cuts applied in the previous analysis are not the optimal ones to improve the significance of the signal with the samples at our disposal. The choice of restricting the analysis to the two E_t-leading jets, implementing only for them the b-tag reconstruction, limits the possibility of increasing the S / B ratio with further kinematical restrictions. The different approch we suggest for future studies consists in selecting events with two b-tagged jets, requesting a minimum energy for both of them without restrict the analysis to the first two E_t-leading jets. This change of perspective is motivated by the fact that, as shown from our preliminary studies, the two b-tagged are more likely to correspond to the two E_t-leading ones in QCD $b\bar{b}$ events (background) rather than in $Z \to b\bar{b}$ (signal); so, looking only at the leading jets pair potentially prevents us from optimizing the signal over background ratio. With this alternative method we end up considering also events in which a boosted Z is produced and its associated jets are not the most energetic in the final state; to fully explore this phase-space region, we have to take into account events with a third well-reconstructed jet, in particular the most energetic jet different from two b-tagged ones. Considering 3 jet-events complicates the study of the QCD contribution, introducing additional jets coming from initial state radiation

and other production processes like flavour excitation and gluon splitting. The idea is to find to appropriate set of kinematical cuts to increase the S / B ratio exploiting the feature of the non b-tagged jet, the "third" one, in order to really unfold the kinematical differences between the two physical processes contributing to the overall signal, i.e. $Z \to b\bar{b}$ and QCD $b\bar{b}$. The first step to built this method of analysis is generate a MC simulation for the QCD process which correctly reproduces all the primary contributions to bottom quark production in hadron colliders, such as flavour creation, flavour excitation and parton shower/fragmentation. The *3-jets* analysis will be the object of further studies in order to built a rigorous procedure for the extraction of the Z signal from the DIJET_BTAG data sample and the improvement of the results presented in [5].

References

[1] U. Grundler, A. Taffard and X. Zhang, *High-P_t muons recommended cuts and efficiencies for Summer 2006*, CDF note 8262 (2006)

[2] B. J. Kim et al., *Z^0 and high mass Drell-Yan cross section measurement in dimuon channel*, CDF note 3510 (1996)

[3] D. Acosta et al. (CDF Collaboration), Phys. Rev. D71, 032001 (2005)

[4] S. Amerio et al., *Triggering on B-jets at CDF II*, IEEE Trans. Nucl. Sci. 56:1690-1695 (2009)

[5] J. Donini et al., *Energy calibration of b-quark jets with $Z \to b\bar{b}$ decays at the Tevatron collider*, `http://arxiv.org/pdf/0801.3906` (2008)

[6] A.Bhatti et al., Nucl. Instrum. Methods, A566, 2 (2006)

[7] M. Dasgupta, L. Magnea, G. Salam, *Non-perturbative QCD effects in jets at hadron colliders*, `http://arxiv.org/pdf/0712.3014` (2007)

[8] D. Acosta et al. (CDF Collaboration), Phys. Rev. D71, 052003 (2005)